

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/41664 A2

(51) International Patent Classification⁷: **H04R 7/04**

Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **COLMENAREZ, Antonio**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(21) International Application Number: PCT/EP01/13304

(22) International Filing Date:
14 November 2001 (14.11.2001)

(74) Agent: **SCHOENMAKER, Maarten**; Internationaal Octrooibureau B.V., Prof Holstlaan 6, NL-5656 AA Eindhoven (NL).

(25) Filing Language: English

(81) Designated State (*national*): JP.

(26) Publication Language: English

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

(30) Priority Data:
09/713,898 16 November 2000 (16.11.2000) US

Published:

— *without international search report and to be republished upon receipt of that report*

(71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors: **TRAJKOVIC, Miroslav**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). **GUTTA, Srinivas**;



WO 02/41664 A2

(54) Title: AUTOMATICALLY ADJUSTING AUDIO SYSTEM

(57) Abstract: An audio generating system that outputs audio through two or more speakers. The audio output of each of the two or more speakers is adjustable based upon the position of a user with respect to the location of the two or more speakers. The system includes at least one image capturing device (such as a video camera) that is trainable on a listening region and coupled to a processing section having image recognition software. The processing section uses the image recognition software to identify the user in an image generated by the image capturing device. The processing section also has software that generates at least one measurement of the position of the user based upon the position of the user in the image.

Automatically adjusting audio system

FIELD OF THE INVENTION

The invention relates to audio systems, such as stereo systems, television audio systems and home theater systems. In particular, the invention relates to systems and methods for adjusting audio systems.

5

BACKGROUND OF THE INVENTION

Particular systems for adjusting the output of various audio systems based on the position of a listener ("user") are known. For example, UK Patent Application GB 2,228,324 describes a system that adjusts the balance of a stereo system as a user moves, in order to maintain the stereo effect for the listener. A signal emitter carried by the user emits signals to two separate receivers that are adjacent to two stereo speakers. The signal emitted may be an ultrasonic signal, infra-red signal or radio signal and may be emitted in response to an initiating signal. (It may also be a wired electrical signal.) The system uses the time it takes a respective receiver (adjacent a speaker) to receive the signal from the signal emitter to determine the distance between the user and the speaker. A distance between the user and each of the two speakers is so calculated. Based on the principle that sound intensity decreases with the cube of the distance from a source, the system uses the distance between each speaker and the user to adjust each speaker so that substantially equal sound intensities are presented to the user from each speaker.

GB 2,228,324 refers to the system determining the position of the user by determining the point where the user's distance from each speaker overlaps, but notes that determining position is not necessary for adjusting stereo balance.

Japanese Patent Abstract 5-137200 detects the position of a viewer in one of five angular zones with respect to the front of a television by pointing a separate infra-red detector at each zone. The balance of the stereo speakers flanking the television screen is said to be adjusted based on the zone the viewer is in.

Japanese Patent Abstract 4-130900 uses elapsed time of light transmission to calculate the distances between a listener and two light emitting and detecting parts. The

distances between the user and the two parts and the distance between the two parts is used to calculate the position of the listener and to adjust the balance of the audio signal.

Similarly, Japanese Patent Abstract 7-302210 uses an infra-red signal to measure the distance between a listening position and a series of speaker and to adjust an appropriate delay time for each speaker based on the distance between the speaker and the listening position.

SUMMARY OF THE INVENTION

One obvious difficulty with the prior art systems is that they either require a user to wear or carry a signal emitter (as in GB 2,228,324) in order to enjoy automatic adjustment of a balance of a stereo system, or, if not, to rely on sensors (such as infra-red sensors) that are unreliable and/or crude in detecting the position of a listener. For example, use of infra-red detectors may fail to detect the listener, resulting in the above-mentioned systems failing to balance properly for the user's position. Moreover, other people (or other items, such as pets) may be sensed by the sensors, resulting in an adjustment in the balance to someone or something other than the listener.

In addition, the above-mentioned systems are not well suited for audio systems more complex than a simple stereo system, for example, a home theater system. A home theater system typically has a multiplicity of speakers positioned about a room that are used to project audio, including audio effects, to a listener. The audio is not simply "balanced" between speakers. Rather, the output of a particular speaker location may be raised and lowered or otherwise coordinated based on the audio effect to be projected to the listener at his or her location. For example, two speakers of a multiplicity of speakers may be driven in phase or out of phase, in order to project a particular audio effect to a listener at the listener's position.

Thus, an accurate determination of the location of each of a multiplicity of speakers with respect to the position of the listener is highly important to certain entertainment experiences. In addition, in order to adjust the required output of a multiplicity of speakers to a changed or changing position of a listener, a more reliable and accurate determination of the listener's position is needed.

Accordingly, the invention provides an audio system (including an audiovisual system) that can automatically adjust to the position of the listener or user of the system, including a change in position of the user. The system uses image capturing and recognition that recognizes some or part of the contours of a human body, i.e., the user. Based on the

position of the user in the field of view, the system determines position information of the user. In one embodiment of the system, for example, the angular position of the user is determined based on the location of the image of the user in the field of view of an imaging capturing device, and the system may adjust the output of two or more speakers based on the determined angle.

The image capturing device may be, for example, a video camera connected to a control unit or CPU that has image recognition software programmed to recognize all or part of the shape of a human body. Various methods of detecting and tracking active contours such as the human body have been developed. For example, a "person finder" that finds and follows people's bodies (or head or hands, for example) in a video image is described in "Pfinder: Real-Time Tracking Of the Human Body" by Wren et al., M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 353, published in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp 780-85 (July 1997), the contents of which are hereby incorporated by reference. Detection of a person (a pedestrian) within an image using a template matching approach is described in "Pedestrian Detection From A Moving Vehicle" by D.M. Gavrila (Image Understanding Systems, DaimlerChrysler Research), Proceedings of the European Conference on Computer Vision, 2000 (available at www.gravila.net), the contents of which are hereby incorporated by reference. Use of a statistical sampling algorithm for detection of a static object in an image and a stochastic model for detection of object motion is described in "Condensation - Conditional Density Propagation For Visual Tracking" by Isard and Black (Oxford Univ. Dept. of Engineering Science), Int. J. Computer Vision, vol. 29, 1998 (available at www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/ISARD1/condensation.html, along with the "Condensation" source code), the contents of which are hereby incorporated by reference. Alternatively, the control unit or CPU may be programmed to recognize the contours of a human head or even the contours of a particular user's face. Software that can recognize faces in images (including digital images) is commercially available, such as the "FaceIt" software sold by Visionics and described at www.faceit.com. Software incorporating such algorithms which may be used to detect human bodies, faces, etc. will be generally referred to as image recognition software, image recognition algorithm and the like in the description below. The position of the recognized body or head relative to the field of view of the camera may be used, for example, to determine the angle of the user's location with respect to the camera. The determined angle may be used to balance or otherwise adjust the audio output and effects to be projected by each speaker to the user's location.

The use of an image capturing device and related image sensing software that identifies the contour of a human body or a particular face makes the detection of the user more accurate and reliable.

Two or more such programmed image capturing devices having overlapping fields of view may be used to accurately determine the location of the user. For example, two separate cameras as described above may be separately located and each may be used to determine the user's position in a reference coordinate system. The user's location may be used by the audio system, for example, to determine the distance between the user's present location and the fixed (known) position of each speaker in the reference coordinate system and to make the appropriate adjustments to the speaker output to provide the proper audio mix to the user's location, such as audio effects in a home theater system.

Thus, in general, the invention comprises an audio generating system that outputs audio through two or more speakers. The audio output of each of the two or more speakers is adjustable based upon the position of a user with respect to the positions of the two or more speakers. The system includes at least one image capturing device (such as a video camera) that is trainable on a listening region and coupled to a processing section having image recognition software. The processing section uses the image recognition software to identify the user in an image generated by the image capturing device. The processing section also has software that generates at least one measurement of the position of the user based upon the position of the user in the image.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a perspective view of a home theater system including automatic detection and locating of a user and adjustment of output in accordance with a first embodiment of the invention;

Fig. 1a is a diagram of portions of the control system of the system of Fig. 1;

Fig. 2a is an image that includes an image of a user captured by a first camera of the system of Fig. 1;

Fig. 2b is an image that includes an image of the user captured by a second camera of the system of Fig. 1;

Fig. 3 is a representative view of a stereo system including automatic detection and locating of a user and adjustment of output in accordance with a second embodiment of the invention; and

Fig. 3a is an image that includes an image of the user captured by a camera of the system of Fig. 3.

DETAILED DESCRIPTION

Referring to Fig. 1, a user 10 is shown positioned amongst audio and visual components of a home theater system. The home theater system is comprised of a video display screen 14 and a series of audio speakers 18a-e surrounding the perimeter of a comfortable viewing area for the display screen 14. The system is also comprised of a control unit 22, shown in Fig. 1 positioned atop the display screen 14. Of course, the control unit 22 may be positioned elsewhere or may be incorporated within the display unit 14 itself. The control unit 22, display screen 14 and speakers 18a-e are all electrically connected with electrical wires and connectors. The wires are typically run beneath carpet in a room or within an adjacent wall, so they are not shown in Fig. 1.

The home theater system of Fig. 1 includes electrical components that produce visual output from display screen 14 and corresponding audio output from speakers 18a-e. The audio and video processing for the home theater output typically occurs in the control unit 22, which may include a processor, memory and related processing software. Such control units and related processing components are known and available in various commercial formats. Audio and video input provided to the control unit 22 may come from a television signal, a cable signal, a satellite signal, a DVD and/or a VCR. The control unit 22 processes the input signal and provides appropriate signals to the driving circuitry of the display screen 14, resulting in a video display, and also processes the input signal and provides appropriate driving signals to the speakers 18a-e, as shown in Fig. 1a.

The audio portion of the signal input to the control unit 22 may be a stereophonic signal or may support more complex audio processing, such as audio effects processing by the control unit 22. For example, the control unit 22 may drive speakers 18b, 18c, 18d in an overlapping sequence in order to simulate a car passing by on the right hand portion of the display. The amplitude and phase of each speaker 18b, 18c, 18d is driven based on received audio signal by the control unit 22, as well as the position of the speaker 18b, 18c, 18d relative to the user 10 as stored in the memory of control unit 22.

The control unit 22 may receive and store the positions of the speakers 18a-e and the position of the user 10 with respect to a common reference system, such as the one defined by origin O and unit vectors (x,y,z) in Fig. 1. The x, y and z coordinates of each speaker 18a-e and the user 10 in the reference coordinate system may be physically measured

or otherwise determined and input to the control unit 22. The position of user 10 in Fig. 1 is shown to have coordinates (X_P, Y_P, Z_P) in the reference coordinate system. The reference coordinate system in general may be located in positions other than shown in Fig. 1. (As described further below, the reference coordinate system in Fig. 1 is chosen to be at the location of a camera in order to facilitate automatic location of the user 10 in accordance with the invention.) Once the coordinates of the speakers 18a-e and user 10 in the reference coordinate system are received by the control unit 22, the control unit 22 may alternatively translate the coordinates to an internal reference coordinate system.

The position of the user 10 and the speakers 18a-e in such a common reference coordinate system enables the control unit 10 to determine the position of the user 10 with respect to each speaker 18a-e. (It is well known that subtracting the coordinates of the user 10 from the coordinates of the speaker 18a determines their relative positions in the reference coordinate system.) Software within the control unit 22 electronically adjusts the driving signals for the audio output (such as volume, frequency, phase) of each speaker based upon the received audio signal, as well as the position of the user 10 relative to the speaker. Electronic adjustment of the audio output by the control unit 22 based on the relative positions of the speakers 18a-e with respect to the user 10 is known in the art. Alternatively, the control system may allow the user to manually adjust the audio output of each speaker 18a-e. Such manual controls of the audio components via the control unit 22 is also known in the art. In both cases, input may be provided to the control unit 22 through a remote that wirelessly interfaces with the control unit 22 and projects a menu on the display screen 14, that allows, for example, input of positional data.

The home theater system of Fig. 1 can also automatically identify the user and the user's location in the reference coordinate system. In the description above, the locations of the user 10 and the speakers 18a-e in the reference coordinate system at origin O were presumed to be known by the control unit 22 based, for example, on manual input provided by the user. Where the position of the user 10 is not known or varies, or an automatic detection and determination of the user's location is otherwise desired, the positions of the speakers 18a-e will still normally be known to the control unit 22, since they usually will remain fixed after they are placed. Thus, the positions of the speakers 18a-e in the reference coordinate system are each manually input to the control system 22 during the initial system set-up and generally remained fixed thereafter. (The speaker location may be changed, of course, and a new position(s) may be input, but this does not occur during normal usage of the system.) Once the user's location is automatically determined by the system, as described

in more detail below, the control unit 22 adjust the audio output to each speaker 18a-e based on the relative locations of the user 10 and the speakers 18a-e, as in the case of manual input of positions, as previously described.

In order to automatically detect the presence and, if present, the location of the user 10 in Fig. 1, the system is further comprised of two video cameras 26a, 26b located atop display screen 14 and directed toward the normal viewing area of the display screen 14. Camera 26a is located at the origin O of the common reference coordinate system. As evident from the description below, video cameras 26a, 26b may be positioned at other locations; the reference coordinate system may also be re-positioned to a different location of camera 26a or elsewhere. Video cameras 26a, 26b interface with the control unit 22 and provide it with images captured in the viewing area. Image recognition software is loaded in control unit 22 and is used by a processor therein to process the video images received from the cameras 26a, 26b. The components, including memory, of the control unit 22 used for image recognition may be separate or may be shared with the other functions of the control unit 22, such as those shown in Fig. 1a. Alternatively, the image recognition may take place in a separate unit.

Fig. 2a depicts the image in the field of view of camera 26a on one side of the display screen of Fig. 1. The image of Fig. 2a is transmitted to control unit 22, where it is processed using, for example, known image recognition software loaded therein. An image recognition algorithm may be used to recognize the contours of a human body, such as the user 10. Alternatively, image recognition software may be used that recognizes faces or may be programmed to recognize a particular face or faces, such as the face of user 10.

Once the image recognition software identifies the contour of a human body or a particular face, the control unit 22 is programmed to determine the point P_i' at the center of the user's 10 head in the image and the coordinates (x', y') with respect to the point O_i' in the upper left-hand corner of the image. As seen, the point O_i' in the image of Fig. 2a corresponds approximately to the point $(0, 0, Z_P)$ in the reference coordinate system of Fig. 1.

Similarly, Fig. 2b depicts the image in the field of view of camera 26b on the other side of the display screen of Fig. 1. In like manner, the image of Fig. 2b is transmitted to control unit 22, where it is processed using image recognition software to recognize the user 10 or the image of the user's face. Because camera 26b is located on the other side of the display screen, the image of the user 10 is located in a different part of the field of view compared to Fig. 2a. The control unit determines the point P_i'' at the center of the user's head

10 in the image of Fig. 2b and the coordinates (x'', y'') with respect to the point O_i'' in the upper left-hand corner of the image.

Having identified the positions P_i' and P_i'' of the user 10 in the camera images shown in Figs. 2a and 2b as having image coordinates (x', y') and (x'', y'') , respectively, the coordinates (X_P, Y_P, Z_P) of the position P of the user 10 in the reference coordinate system of Fig. 1 may be uniquely determined using standard techniques of computer vision known as the "stereo problem". Basic stereo techniques of three dimensional computer vision are described for example, in "Introductory Techniques for 3-D Computer Vision" by Trucco and Verri, (Prentice Hall, 1998) and, in particular, Chapter 7 of that text entitled "Stereopsis", the contents of which are hereby incorporated by reference. Using such well-known techniques, the relationship between the user's position P in Fig. 1 (having unknown coordinates (X_P, Y_P, Z_P)) and the image position P_i' of the user in Fig. 2a (having known image coordinates (x', y')) is given by the equations:

$$x' = X_P/Z_P \quad (\text{Eq. 1})$$

$$y' = Y_P/Z_P \quad (\text{Eq. 2})$$

Similarly, the relationship between the user's position P in Fig. 1 and the image position P_i'' of the user in Fig. 2b (having known image coordinates (x'', y'')) is given by the equations:

$$x'' = (X_P - D)/Z_P \quad (\text{Eq. 3})$$

$$y'' = Y_P/Z_P \quad (\text{Eq. 4})$$

where D is the distance between cameras 26a, 26b. One skilled in the art will recognize that the terms given in Eqs. 1-4 are up to linear transformations defined by camera geometry.

Equations 1-4 have three unknown variables (coordinates X_P, Y_P, Z_P), thus the simultaneous solution gives the values of X_P, Y_P , and Z_P and thus gives the position of the user 10 in the reference coordinate system of Fig. 1.

If required, the coordinates (X_P, Y_P, Z_P) may be translated to another internal coordinate system of the control unit 22. The processing required to determine the position (X_P, Y_P, Z_P) of the user and to translate the radial coordinates to another reference coordinate, if necessary, may also take place in a processing unit other than control unit 22. For example, it may take place in a processing unit that also supports the image recognition processing, thus comprising a separate processing unit dedicated to the tasks of image detection and location.

As noted above, the fixed positions of speakers 18a-e are known to the control unit 22 based on prior input. For example, once each speaker 18a-e is placed in the room as shown in Figs. 1, the coordinates (x, y, z) of each speaker 18a-e in the reference coordinate

system, and the distance D between cameras 26a, 26b may be measured and input in memory in the control unit 22. The coordinates (X_p, Y_p, Z_p) of the user 10 as determined using the image recognition software (along with the post-recognition processing of the stereo problem described above) and the pre-stored coordinates of each speaker may then be used to
5 determine the position of the user 10 with respect to each speaker 18a-e. As previously described, the audio processing of the control unit 22 may then appropriately adjust the audio output (including amplitude, frequency and phase) of each speaker 18a-e based upon the input audio signal and the position of the user 10 with respect to the speakers 18a-e.

The use of the video cameras 26a, 26b, image recognition software, and post-
10 recognition processing to determine a detected user's position thus allows the location of the user of the home theater system of Fig. 1 to be automatically detected and determined. If the user moves, the processing is repeated and a new position is determined for the user, and the control unit 22 uses the new location to adjust the audio signals output by speakers 18a-e.

The automatic detection feature may be turned off so that the output of the
15 speakers is based on a default or a manual input of the location of the user 10. The image recognition software may also be programmed to recognize, for example, a number of different faces and the face of a particular user may be selected for recognition and automatic adjustment. Thus, the system may adjust to the position of a particular user in the viewing area. Alternatively, the image recognition software may be used to detect all faces or human
20 bodies in the viewing area and the processing may then automatically determine each of their respective locations. The adjustment of the audio output of each speaker 18a-e may be determined by an algorithm that attempts to optimize the aural experience at the location of each detected user.

Although the embodiment of Fig. 1 depicted a home theater system, the
25 automatic detection and adjustment may be used by other audiovisual systems or other purely audio systems. It may be used, for example, with a stereo system having a number of speakers to adjust the volume at each speaker location based on the determined location of the user with respect to the speakers in order to maintain a proper (or pre-determined) balance of the stereophonic sound at the location of the user.

Thus, a simpler embodiment of the invention applied to a two speaker stereo
30 system is shown in Fig. 3. The basic components of the stereo system comprise a stereo amplifier 130 attached to two speakers 100a, 100b. A camera 110 is used to detect an image of a listening region, including the image of a listener 140 in the listening region. The relative positions of the speakers 100a, 100b, camera 110 and user 140 are shown from

above, or projected into the plane of the floor. Fig. 3 also shows a simple reference coordinate system in the plane, having an origin O at the camera and comprised of the angle of an object with respect to the axis A of the camera 110. Thus, the angle \exists is the angular position of speaker 100a, the angle N is the angular position of speaker 100b and the angle 2 is the angular position of the user 140. (Fig. 3 shows the top of the user's head.)

In the system of fig. 3, the user 140 is assumed to listen to the stereo in the central region of Fig. 3 at an approximate distance D from the origin O. The speakers 100a, 100b have a default balance at the position D along the axis A, which is approximately at the center of the listening area.

The angles \exists and N of the positions of speakers 100a, 100b are measured and pre-stored in processing unit 120. The image captured by the camera 110 is transferred to the processing unit 120 that includes image recognition software that detects the contour of a human body, a particular face, etc., as described in the embodiment above. The location of the detected body or face in the image is used by the processing unit to determine the angle 2 corresponding to the position of the user 140 in the reference coordinate system. For example, referring to Fig. 3a, a first order determination of the angle 2 is:

$$2 = (x/W)(P)$$

where x is the horizontal image distance measured by the processing unit 120 from the center C of the image, W is the total horizontal width of the image and the P is the field of view, or, equivalently, the angular width of the scene, as fixed by the camera.

The processing unit 120 in turn sends a signal to the amplifier that adjusts the balance of speakers 100a, 100b based on the relative angular positions of the user 140 and the speakers 100a, 100b. For example, the output of speaker 110a is adjusted using a factor $(\exists-2)$ and the output of speaker 110b is adjusted using a factor $(N+2)$. Thus, the balance of speakers 100a, 100b is thus automatically adjusted based upon the position of the user 140 with respect to the speakers 100a, 100b. As previously noted, it is assumed in the system of Fig. 4 that the user 140 remains in a central listening region in Fig. 3, at an approximate distance D from the origin O. Thus, the adjustment of the balance is based on the angular position 2 of the user is an acceptable first order adjustment.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, but rather it is intended that the scope of the invention is as defined by the scope of the appended claims.

CLAIMS:

1. An audio generating system that outputs audio through two or more speakers (18a-e, 100a, 100b), the audio output of each of the two or more speakers (18a-e, 100a, 100b) being adjustable based upon the location of a user with respect to the location of the two or more speakers (18a-e, 100a, 100b), the system comprising at least one image capturing
5 device (26a, 26b, 110) trainable on a listening region and coupled to a processing section (22, 120) having image recognition software that identifies the user in an image generated by the image capturing device (26a, 26b, 110), the processing section (22, 120) having additional software that generates at least one measurement of the position of the user based upon the position of the user in the image.

10 2. The audio generating system of Claim 1, wherein the system is part of an audiovisual system.

15 3. The audio generating system of Claim 2, wherein the audiovisual system is a home theater system.

4. The audio generating system of Claim 1, wherein the processing section (22, 120) adjusts the audio output of at least one of the speakers (18a-e, 100a, 100b) based upon the at least one measurement of the position of the user.

20 5 The audio generating system of Claim 4 wherein the processing section (22, 120) is comprised of a single processing unit that identifies the user in the image, generates the at least one measurement of the position of the user and adjusts the audio output of at least one of the speakers (18a-e, 100a, 100b) based upon the at least one measurement of the
25 position of the user.

6. The audio generating system of Claim 4 wherein the processing section is comprised of a first processing unit that identifies the user in the image and generates the at least one measurement of the position of the user and a second processing unit that adjusts

the audio output of at least one of the speakers (18a-e, 100a, 100b) based upon the at least one measurement of the position of the user.

7. The audio generating system of Claim 1, wherein the at least one image capturing device is a video camera (26a, 26b, 110).

8. The audio generating system of Claim 7, wherein the at least one measurement of position of the user is an angle in a reference coordinate system.

9. The audio generating system of Claim 7, wherein the processing section (120) uses the angle to adjust the output of at least one speaker (110a, 110b).

10. The audio generating system of Claim 1, wherein the at least one image capturing device is two or more video cameras (26a, 26b, 110).

11. The audio generating system of Claim 10, wherein the processing section (22) determines a position of the user in a reference coordinate system using the positions of the user in the images generated by each of the two or more video cameras (26a, 26b).

12. The audio generating system of Claim 11, wherein the processing section (22) uses a stereo technique of three dimensional computer vision to determine the position of the user in the reference coordinate system using the positions of the user in the images generated by each of the two or more video cameras (26a, 26b).

13. The audio generating system of Claim 11, wherein the processing section (22) uses the position of the user in the reference coordinate system and the positions of the two or more speakers in the reference coordinate system to determine the distance between the user and each of the two or more speakers (26a, 26b).

14. The audio generating system of Claim 13, wherein the distance between the user and each of the two or more speakers is used to adjust the audio output of at least one of the two or more speakers (26a, 26b).

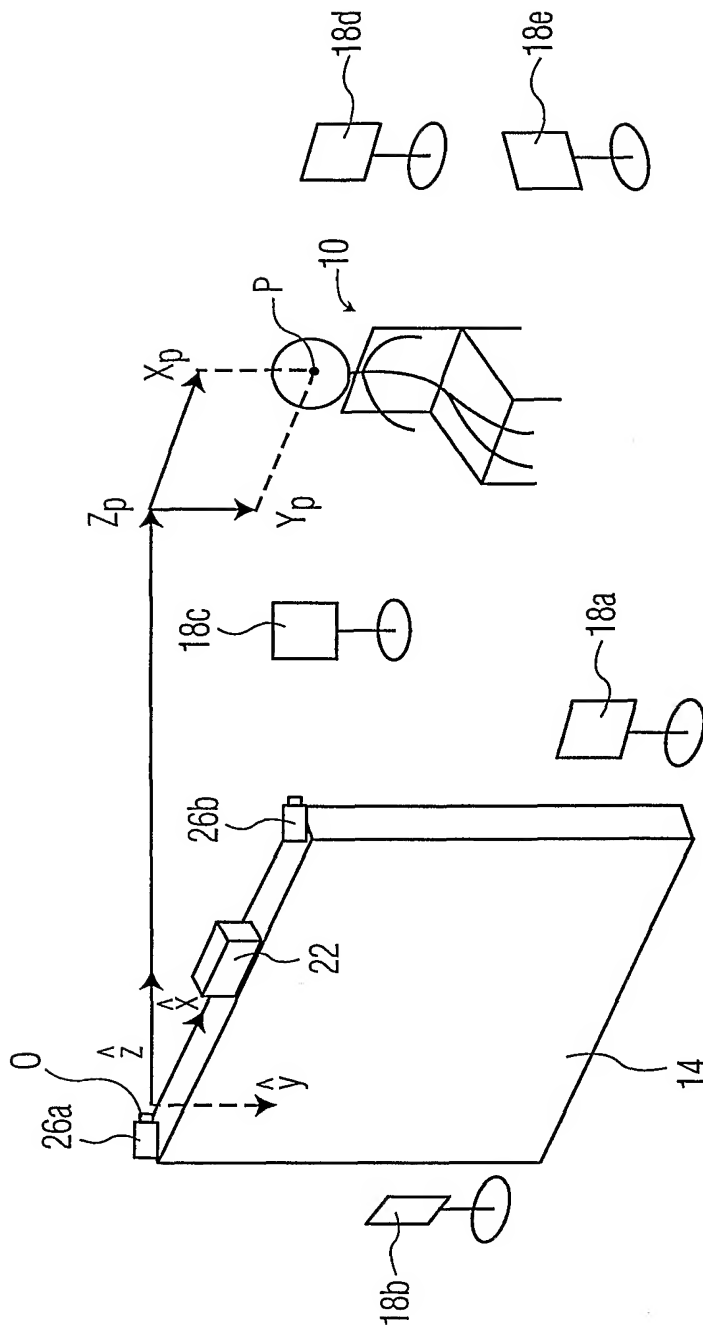


FIG. 1

2/4

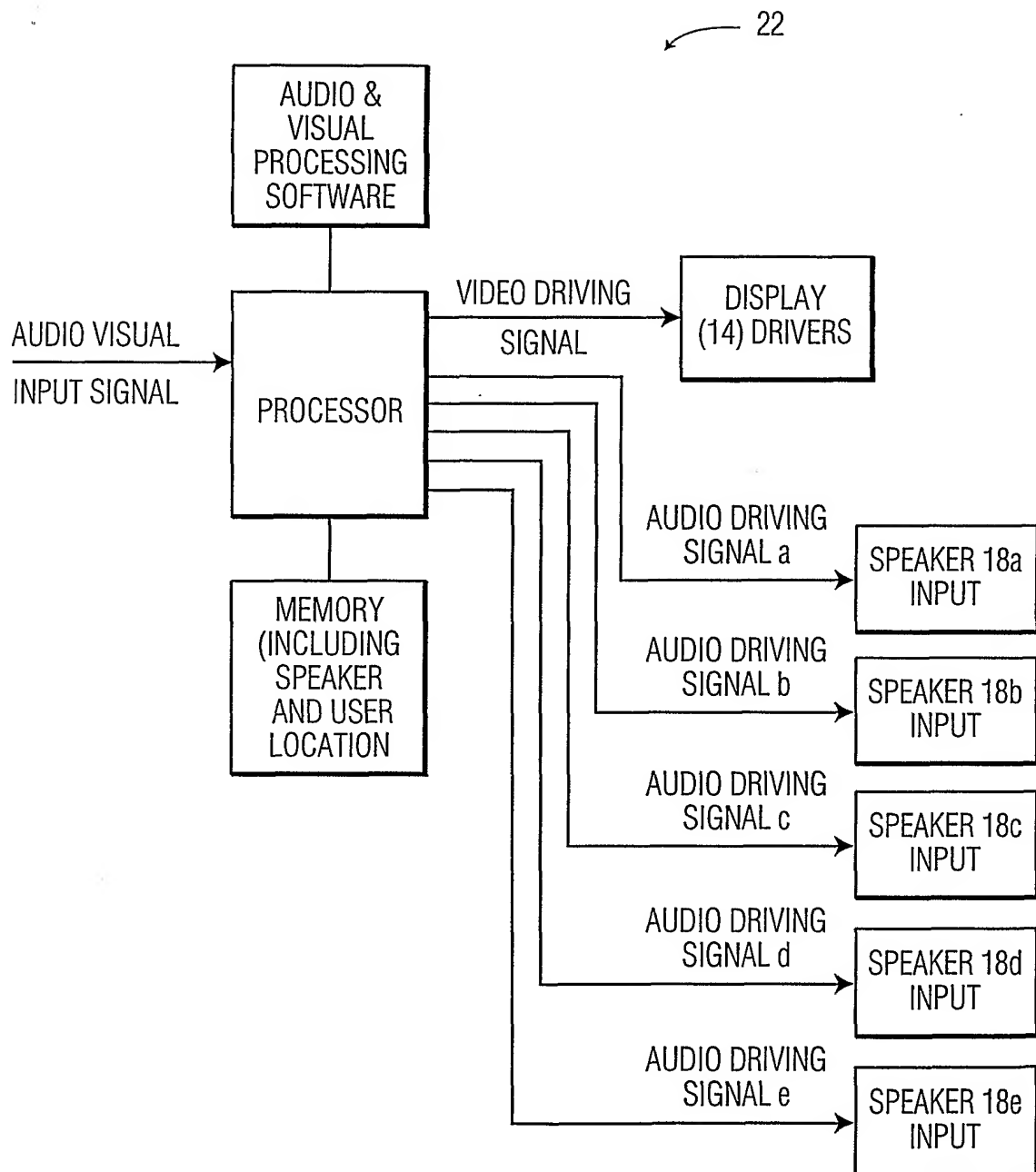


FIG. 1A

3/4

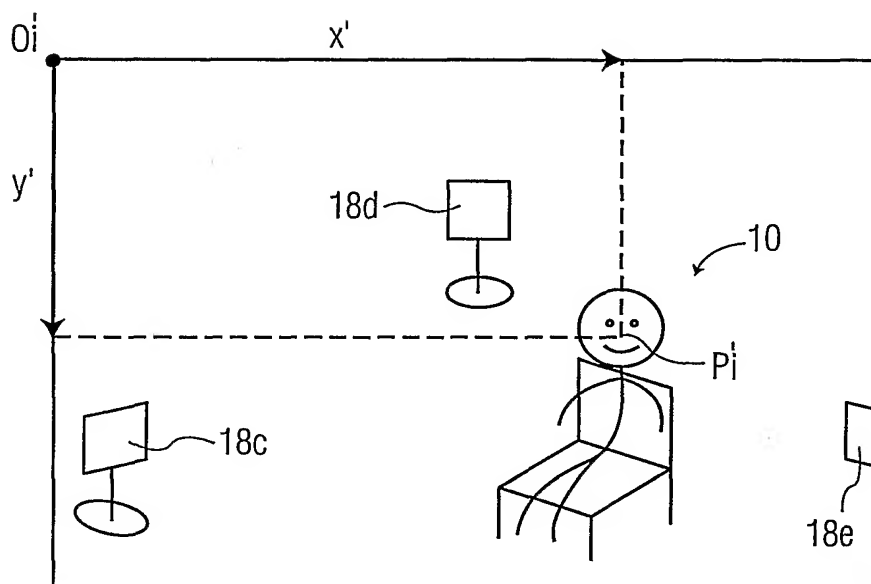


FIG. 2A

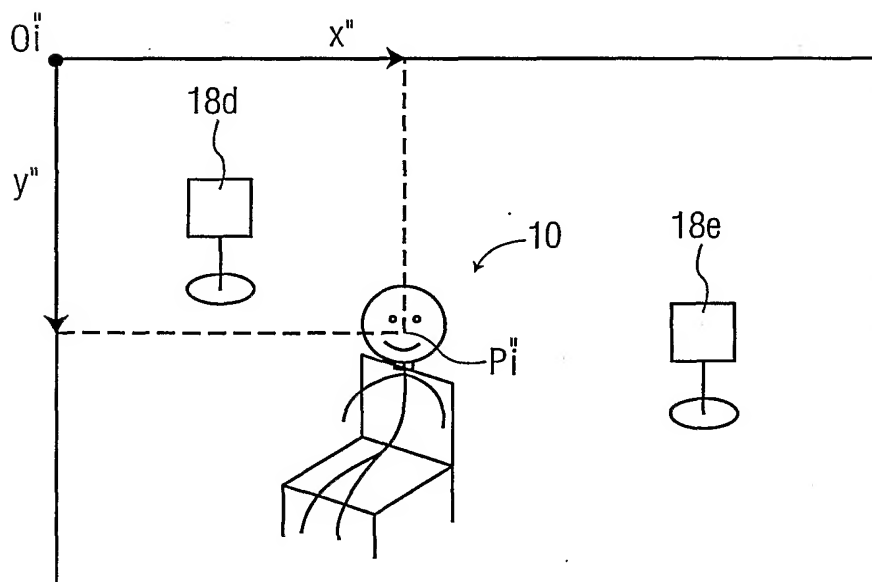


FIG. 2B

4/4

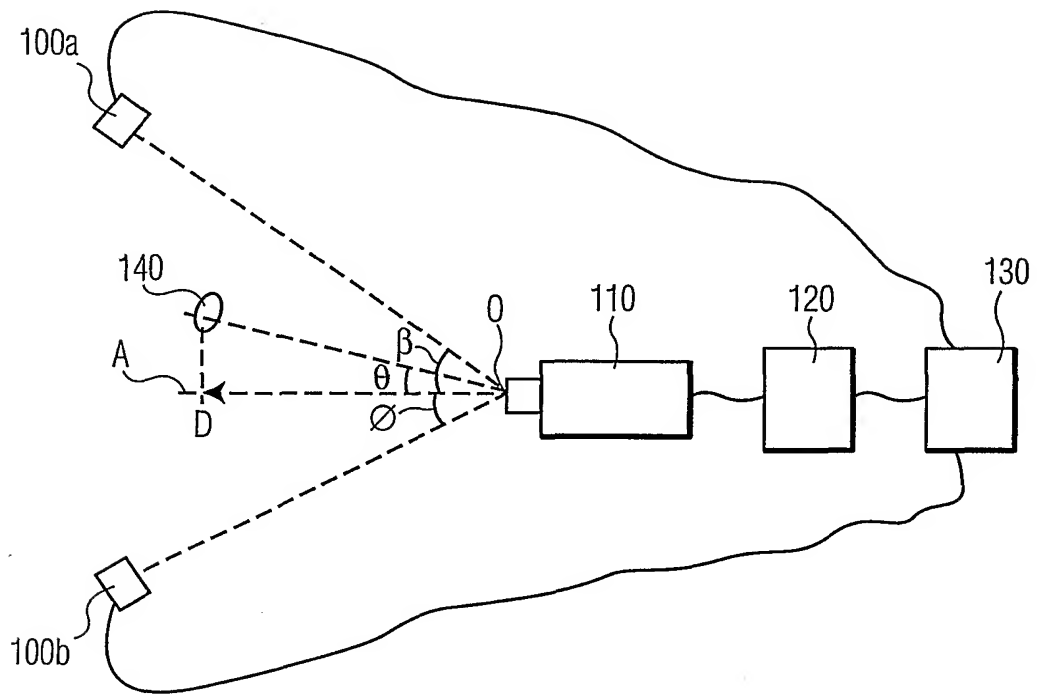


FIG. 3.

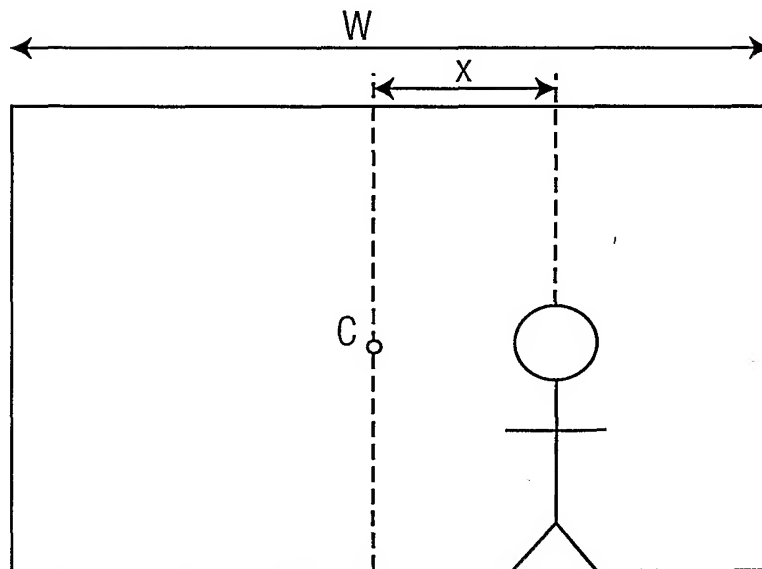


FIG. 3A